



PRIMERA EVALUACIÓN DEL SEGUNDO
BIMESTRE
ANÁLISIS DE DATOS



Análisis visual de datos.

Nombre: Carlos Antonio Ushiña Davila

Realizar el análisis exploratorio visual de acuerdo con el dataset asignado.

Caso 1



El dataset corresponde a 1950 películas hasta el 2023 tiene 11 columnas que se puede ver el detalle en la página de Kaggle.

Fuente:

<https://www.kaggle.com/datasets/kianindeed/imdb-movie-dataset-dec-2023>

Caso 2



El dataset corresponde al ranking de universidades a nivel mundial, el detalle de cada columna se lo puede revisar en la página de Kaggle.

Fuente:



PRIMERA EVALUACIÓN DEL SEGUNDO BIMESTRE ANÁLISIS DE DATOS



<https://www.kaggle.com/datasets/raymondtoo/the-world-university-rankings-2016-2024>

Caso 3

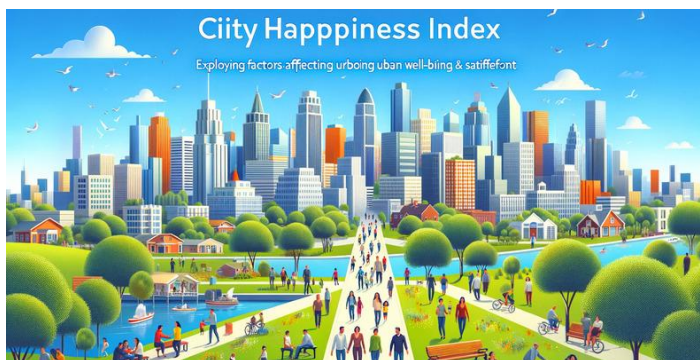


El dataset corresponde al registro de vehículos eléctricos, el detalle de cada columna se lo puede revisar en la página de Kaggle.

Fuente:

<https://www.kaggle.com/datasets/yashusinghal/electric-vehicle-population-dataset>

Caso 4



El dataset corresponde al índice de felicidad en algunas ciudades y los factores que apalancan ese índice, el detalle de cada columna se lo puede revisar en la página de Kaggle.

Fuente:

<https://www.kaggle.com/datasets/emirhanai/city-happiness-index-2024>



PRIMERA EVALUACIÓN DEL SEGUNDO BIMESTRE ANÁLISIS DE DATOS



Instrucciones

- 1.- Generar 10 visualizaciones utilizando cualquier librería.
- 2.- De cada visualización generar 2 conclusiones
- 3.- Destacar 2 conclusiones y sustentar con datos externos, comparar que dicen otros sitios web acerca del caso de estudio respectivo, al menos un párrafo con fuentes de contraste.

PROCEDIMIENTO

Primer paso voy con el código para poder saber que tipo de nombres tengo en la columna para poder implementarlos y ocuparlos:

En mi caso voy a implementar la misma opción de obtener la información, pero presentada en diferente gráfica, con el fin de cumplir con el primer punto donde me pide realizar 10 graficas donde se visualicen los datos entre para notar la diferencia de formas que hay entre ellos.

```
[11]: import pandas as pd

# Cargar los datos desde el archivo CSV
data = pd.read_csv(r"C:\Users\ANALISIS DE DATOS\Desktop\Prueba_Analisis_CarlosUshiña\Electric_Vehicle_Population_Data.csv")

# Obtener los nombres de las columnas
columnas = data.columns
print(columnas)

Index(['VIN (1-10)', 'County', 'City', 'State', 'Postal Code', 'Model Year',
      'Make', 'Model', 'Electric Vehicle Type',
      'Clean Alternative Fuel Vehicle (CAFV) Eligibility', 'Electric Range',
      'Base MSRP', 'Legislative District', 'DOL Vehicle ID',
      'Vehicle Location', 'Electric Utility', '2020 Census Tract'],
      dtype='object')
```

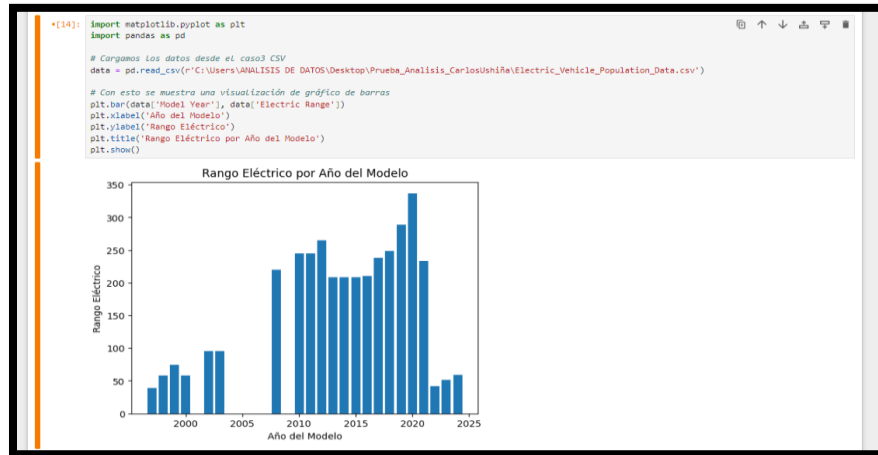
Una vez obtenido los nombres procedemos con la respectiva ocupación para tener la gráfica que nosotros deseemos:



PRIMERA EVALUACIÓN DEL SEGUNDO BIMESTRE ANÁLISIS DE DATOS



1. En este caso escogí el diagrama de barras simple donde me ayuda con el tipo de visualización.

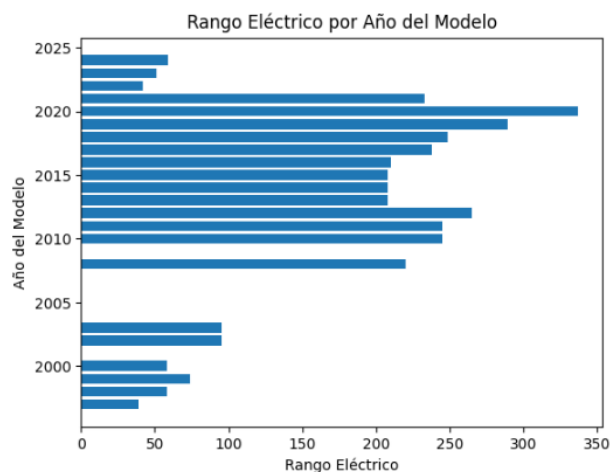


2. En este caso escogí el mismo dato pero que se invierta por decirlo de alguna manera y me muestre los diagramas en horizontal para tener una mejor apreciación del código:

```
[15]: import matplotlib.pyplot as plt
import pandas as pd

# Obtenemos los datos desde el archivo CSV
data = pd.read_csv(r"C:\Users\ANALISIS DE DATOS\Desktop\Prueba_Analisis_CarlosUshiña\Electric_Vehicle_Population_Data.csv")

# con esto creamos una visualización de gráfico de barras horizontales
plt.barh(data['Model Year'], data['Electric Range'])
plt.xlabel('Rango Eléctrico')
plt.ylabel('Año del Modelo')
plt.title('Rango Eléctrico por Año del Modelo')
plt.show()
```

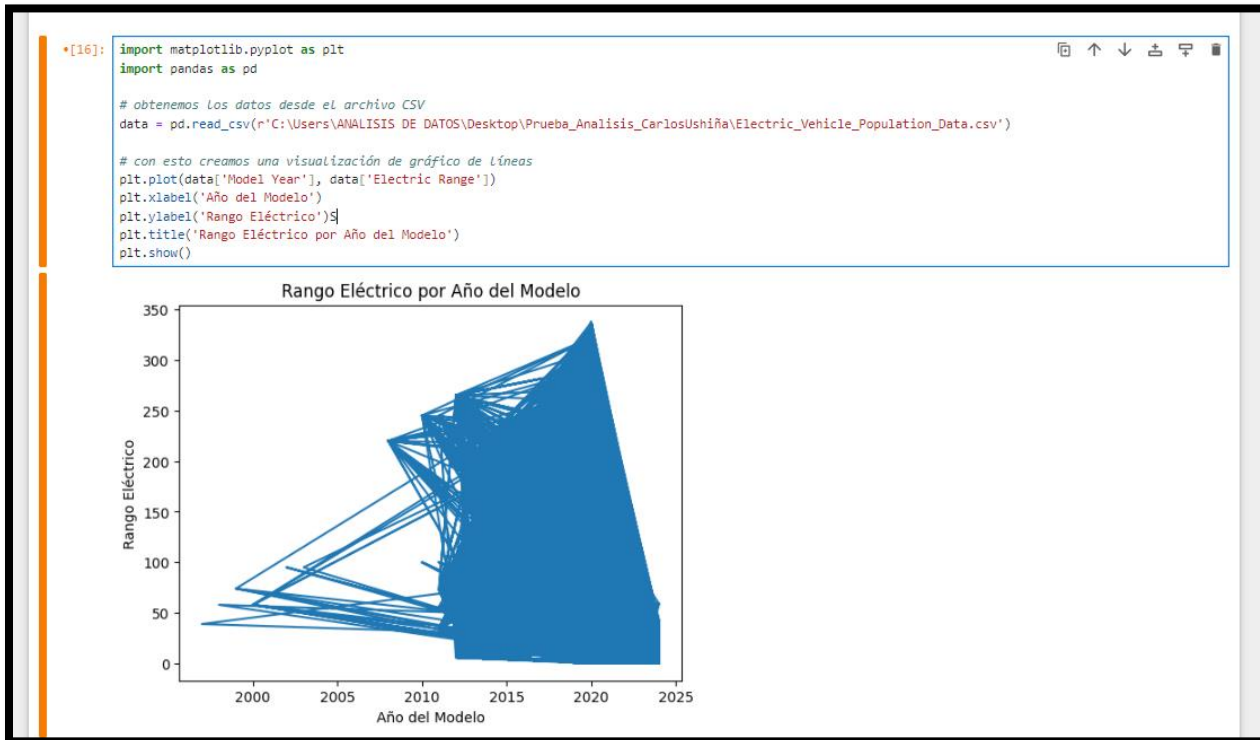




PRIMERA EVALUACIÓN DEL SEGUNDO BIMESTRE ANÁLISIS DE DATOS



3. Con este código obtenemos la información en la presentación de líneas.



4. Cambiando los datos podemos observar otro estilo donde vamos a practicar con graficas 3d donde tenemos que ocupar una librería 3d



PRIMERA EVALUACIÓN DEL SEGUNDO BIMESTRE ANÁLISIS DE DATOS



```
[18]: import matplotlib.pyplot as plt
import pandas as pd
from mpl_toolkits.mplot3d import Axes3D # Importamos la herramienta para gráficos 3D

# Obtenemos los datos desde el archivo CSV
data = pd.read_csv(r'C:\Users\ANALISIS DE DATOS\Desktop\Prueba_Analisis_CarlosUshña\Electric_Vehicle_Population_Data.csv')

# Creamos una visualización de gráfico de columnas 3D
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d') # Agregar un eje 3D al gráfico

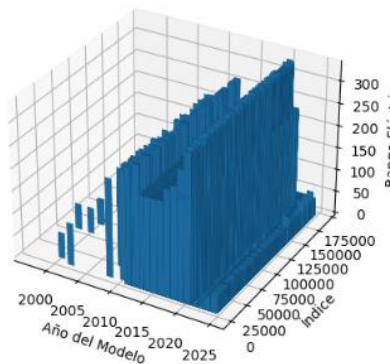
# Configuramos los datos para el gráfico de columnas 3D
x = data['Model Year']
y = range(len(data['Model Year']))
z = data['Electric Range']

# Crear el gráfico de columnas 3D
ax.bar3d(x, y, 0, 1, 1, z)

# Configuramos etiquetas y título
ax.set_xlabel('Año del Modelo')
ax.set_ylabel('Índice')
ax.set_zlabel('Rango Eléctrico')
plt.title('Rango Eléctrico por Año del Modelo')

# Mostramos el gráfico
plt.show()
```

Rango Eléctrico por Año del Modelo



5. En este paso ejecutamos el gráfico cónico para tener un mayor control sobre lo que se presenta en el archivo csv. Los respectivos comentarios se encuentran en el documento ipynb.



PRIMERA EVALUACIÓN DEL SEGUNDO BIMESTRE ANÁLISIS DE DATOS



```
[20]: import matplotlib.pyplot as plt
import pandas as pd

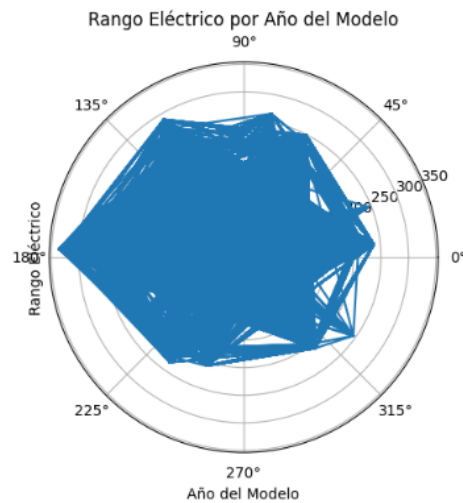
# obtenemos los datos desde el archivo CSV
data = pd.read_csv(r'C:\Users\ANALISIS DE DATOS\Desktop\Prueba_Analisis_CarlosUshiña\Electric_Vehicle_Population_Data.csv')

# creamos una figura y añadimos un gráfico cónico
fig = plt.figure()
ax = fig.add_subplot(111, projection='polar')

# graficamos los datos en el gráfico cónico
ax.plot(data['Model Year'], data['Electric Range'])

# personalizamos el gráfico
ax.set_xlabel('Año del Modelo')
ax.set_ylabel('Rango Eléctrico')
ax.set_title('Rango Eléctrico por Año del Modelo')

# mostramos el gráfico
plt.show()
```



6. En este caso vemos la grafica de diagrama de pastel en 3d para poder tener una visualización mas clara de los modelos y rangos eléctricos.



PRIMERA EVALUACIÓN DEL SEGUNDO BIMESTRE ANÁLISIS DE DATOS



```
[21]: import matplotlib.pyplot as plt
import pandas as pd

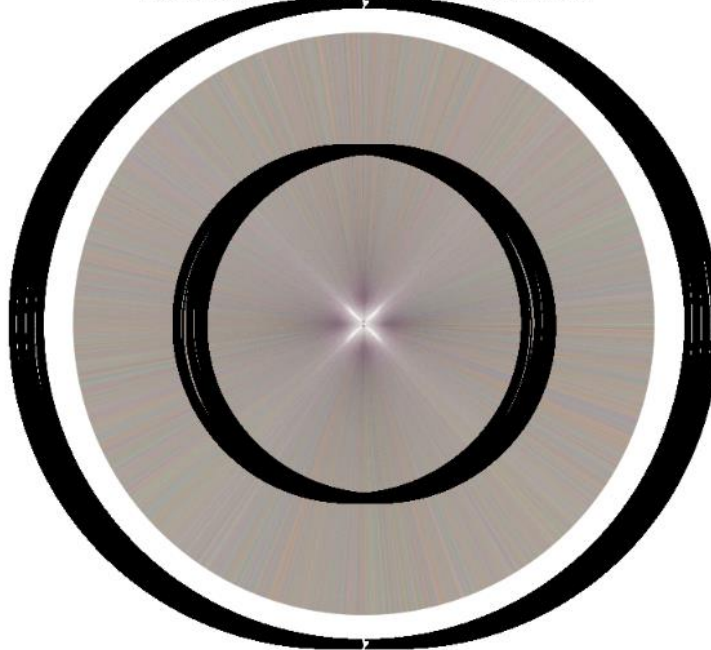
# obtenemos los datos desde el archivo CSV
data = pd.read_csv(r"C:\Users\ANALISIS DE DATOS\Desktop\Prueba_Analisis_CarlosUshina\Electric_Vehicle_Population_Data.csv")

# creamos un gráfico de pastel
plt.figure(figsize=(8,8))
plt.pie(data['Electric Range'], labels=data['Model Year'], autopct='%1.1f%%', startangle=140)
plt.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.

# añadimos título
plt.title('Distribución del Rango Eléctrico por Año del Modelo')

# mostramos el gráfico
plt.show()
```

Distribución del Rango Eléctrico por Año del Modelo



7. A partir del gráfico de dispersión, se observa que no existe una clara tendencia de aumento o disminución del rango eléctrico en función del año del modelo.



PRIMERA EVALUACIÓN DEL SEGUNDO BIMESTRE ANÁLISIS DE DATOS



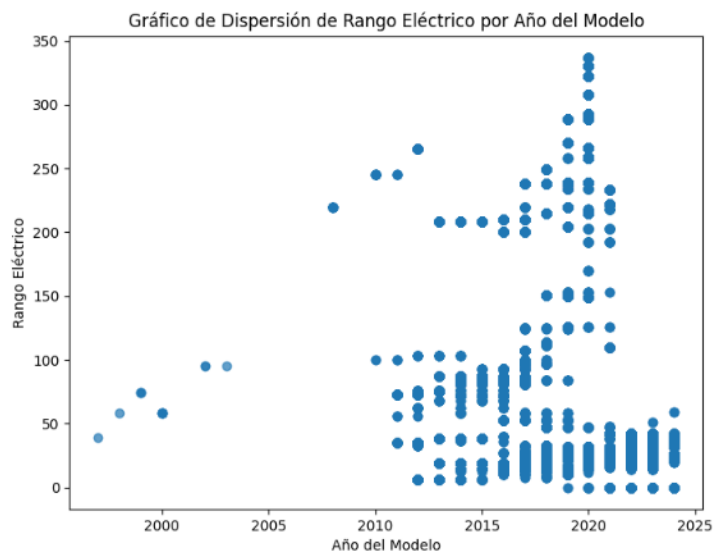
```
[22]: import matplotlib.pyplot as plt
import pandas as pd

# obtenemos los datos desde el archivo CSV
data = pd.read_csv(r'C:\Users\ANALISIS DE DATOS\Desktop\Prueba_Analisis_CarlosUshiña\Electric_Vehicle_Population_Data.csv')

# creamos un gráfico de dispersión
plt.figure(figsize=(8, 6))
plt.scatter(data['Model Year'], data['Electric Range'], alpha=0.7) # alpha controla la transparencia de los puntos

# añadimos etiquetas y título
plt.xlabel('Año del Modelo')
plt.ylabel('Rango Eléctrico')
plt.title('Gráfico de Dispersión de Rango Eléctrico por Año del Modelo')

# mostramos el gráfico
plt.show()
```



8. A partir del histograma, se puede observar la distribución de frecuencia del rango eléctrico de los vehículos



PRIMERA EVALUACIÓN DEL SEGUNDO BIMESTRE ANÁLISIS DE DATOS



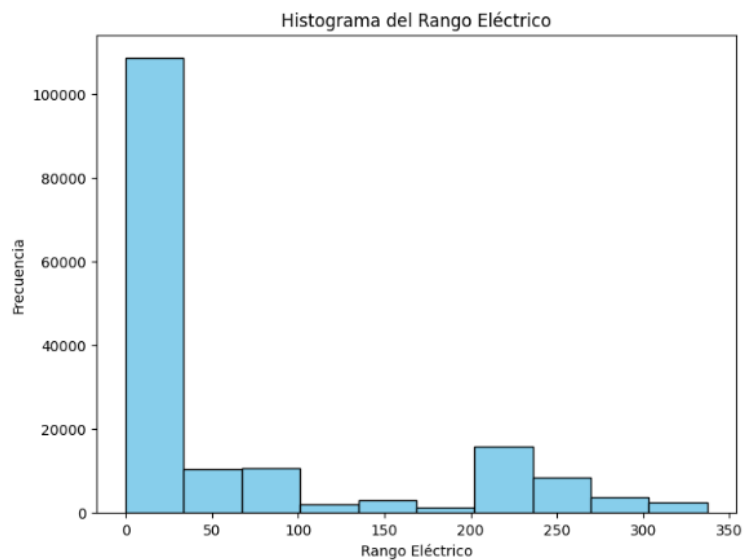
```
[23]: import matplotlib.pyplot as plt
import pandas as pd

# obtenemos los datos desde el archivo CSV
data = pd.read_csv(r'C:\Users\ANALISIS DE DATOS\Desktop\Prueba_Analisis_CarlosUshiña\Electric_Vehicle_Population_Data.csv')

# creamos un gráfico de histograma
plt.figure(figsize=(8, 6))
plt.hist(data['Electric Range'], bins=10, color='skyblue', edgecolor='black') # bins define el número de barras en el histograma

# añadimos etiquetas y título
plt.xlabel('Rango Eléctrico')
plt.ylabel('Frecuencia')
plt.title('Histograma del Rango Eléctrico')

# mostramos el gráfico
plt.show()
```



9. El gráfico piramidal proporciona una representación visual única de la distribución del rango eléctrico en función del año del modelo



PRIMERA EVALUACIÓN DEL SEGUNDO BIMESTRE ANÁLISIS DE DATOS



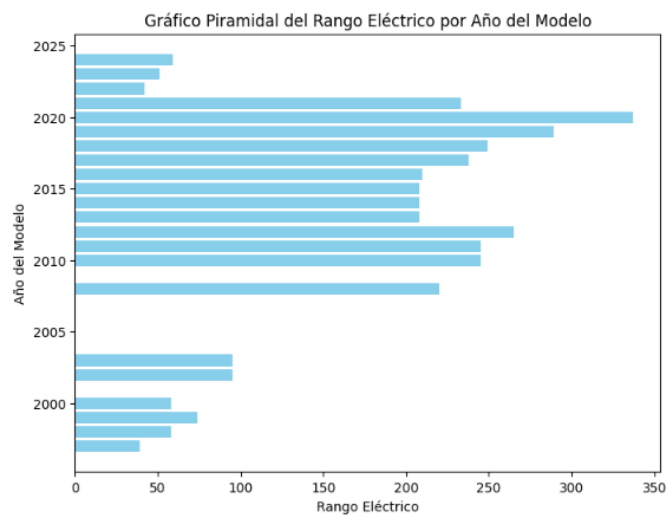
```
[24]: import matplotlib.pyplot as plt
import pandas as pd

# obtenemos los datos desde el archivo CSV
data = pd.read_csv(r"C:\Users\ANALISIS DE DATOS\Desktop\Prueba_Analisis_CarlosUshiña\Electric_Vehicle_Population_Data.csv")

# creamos un gráfico piramidal
plt.figure(figsize=(8, 6))
plt.barh(data["Model Year"], data["Electric Range"], color='skyblue')

# añadimos etiquetas y título
plt.xlabel('Rango Eléctrico')
plt.ylabel('Año del Modelo')
plt.title('Gráfico Piramidal del Rango Eléctrico por Año del Modelo')

# mostramos el gráfico
plt.show()
```



10. El gráfico de líneas representa la evolución del rango eléctrico de los vehículos eléctricos a lo largo de los años del modelo. Muestra de manera clara y visual la tendencia general del rango eléctrico en relación con el año del modelo

Si bien es similar a la gráfica de pastel, es diferente en la estructura como tal, pero debido a la extensa cantidad de datos, se visualizan casi de la misma manera



PRIMERA EVALUACIÓN DEL SEGUNDO BIMESTRE ANÁLISIS DE DATOS



```
[26]: import matplotlib.pyplot as plt
import pandas as pd

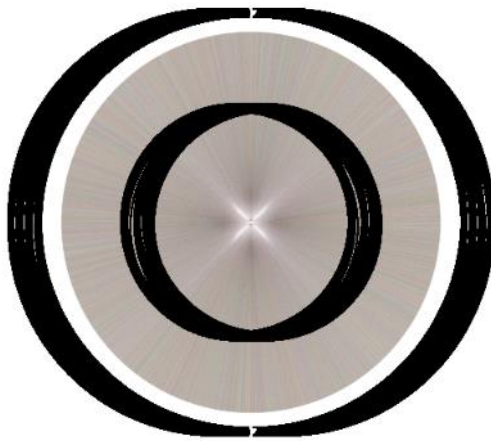
# obtenemos los datos desde el archivo CSV
data = pd.read_csv(r'C:\Users\ANALISIS DE DATOS\Desktop\Prueba_Analisis_CarlosUshina\Electric_Vehicle_Population_Data.csv')

# creamos un gráfico circular
plt.figure(figsize=(8, 6))
plt.pie(data['Electric Range'], labels=data['Model Year'], autopct='%1.1f%%', startangle=140, colors=plt.cm.tab20.colors)

# añadimos título
plt.title('Gráfico Circular del Rango Eléctrico por Año del Modelo')

# mostramos el gráfico
plt.show()
```

Gráfico Circular del Rango Eléctrico por Año del Modelo



Por un ejemplo al realizar un gráfico de área, es similar al gráfico de líneas pero con sus diferentes perspectivas, por ejemplo:



PRIMERA EVALUACIÓN DEL SEGUNDO BIMESTRE ANÁLISIS DE DATOS



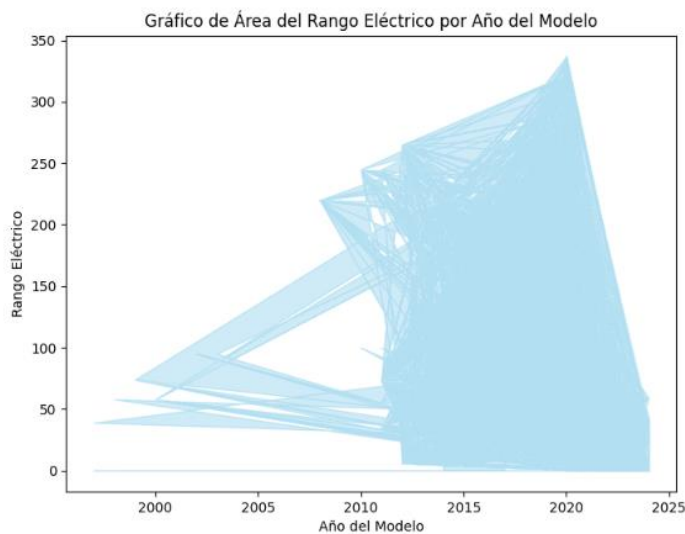
```
[27]: import matplotlib.pyplot as plt
import pandas as pd

# obtenemos los datos desde el archivo CSV
data = pd.read_csv(r'C:\Users\ANALISIS DE DATOS\Desktop\Prueba_Analisis_CarlosUshiña\Electric_Vehicle_Population_Data.csv')

# creamos un gráfico de área
plt.figure(figsize=(8, 6))
plt.fill_between(data['Model Year'], data['Electric Range'], color='skyblue', alpha=0.4)

# añadimos etiquetas y título
plt.xlabel('Año del Modelo')
plt.ylabel('Rango Eléctrico')
plt.title('Gráfico de Área del Rango Eléctrico por Año del Modelo')

# mostramos el gráfico
plt.show()
```



[]:

2. Comentarios de cada grafica ocupada:

Los comentarios de cada grafica se encuentran en el archivo.ipynb, pero aquí aprovechare para dar unos comentarios de cada grafico que ocupe y para que nos serian mas útiles.

- I. Gráfico de barra simple: Este tipo de gráfico es efectivo para comparar cantidades grandes como en el caso 3 entre diferentes categorías. Es útil para visualizar datos discretos y permite una fácil comparación entre las categorías.



PRIMERA EVALUACIÓN DEL SEGUNDO BIMESTRE ANÁLISIS DE DATOS



- II. Gráfico de barras horizontales: Las barras horizontales son útiles cuando se tienen etiquetas largas o cuando se quiere enfatizar la diferencia en longitud entre las barras. Este tipo de gráfico es ideal para comparar cantidades entre diferentes categorías con etiquetas largas.
- III. Gráfico de líneas: Los gráficos de líneas son útiles para mostrar tendencias a lo largo del tiempo o en un eje continuo si se tiene en consideración que el archivo es extenso. Son efectivos para visualizar cambios y patrones en los datos a lo largo de una escala continua.
- IV. Gráfico de columnas 3D: Los gráficos de columnas 3D añaden una dimensión adicional a los gráficos de columnas estándar, lo que puede hacer que la visualización sea más compleja. A menudo, los gráficos 3D pueden dificultar la comparación precisa de las alturas de las columnas, pero en este caso casi no se puede apreciar debido a la magnitud de los campos y los datos en general.
- V. Gráfica cónica: Este tipo de gráfica representa datos en forma de cono, con la base del cono representando la magnitud de los datos. Es menos común en comparación con otros tipos de gráficos, pero aún no se tiene una buena forma con respecto al documento por lo mencionado.
- VI. Gráfica de pastel con todos los datos: La gráfica de pastel es ideal para mostrar partes de un todo. Sin embargo, al representar muchos datos en un solo pastel, puede volverse confuso y difícil de interpretar.
- VII. Gráfica de dispersión: Es útil para visualizar la relación entre dos conjuntos de datos numéricos. Permite identificar patrones, tendencias y posibles correlaciones entre las variables.
- VIII. Gráfica de histograma: Los histogramas son útiles para visualizar la distribución de datos y la frecuencia con la que ocurren ciertos valores dentro de un conjunto de datos.
- IX. Gráfica piramidal: Este tipo de gráfica es menos común, pero puede ser utilizado para visualizar datos de manera piramidal, donde las categorías se presentan en forma de pirámide.
- X. Gráfica circular: Las gráficas circulares son efectivas para mostrar proporciones de un todo. Son útiles para representar categorías y sus porcentajes en relación con el total, aunque sea similar a la gráfica de pastel tienen sus pequeñas diferencias, pero debido a la magnitud del documento, no se puede apreciar bien la misma.

3. Destacar 2 conclusiones y sustentar con datos externos, comparar que dicen otros sitios web acerca del caso de estudio respectivo, al menos un párrafo con fuentes de contraste.

- ✓ Los gráficos ocupados nos muestran una tendencia general al alza en el rango eléctrico de los vehículos eléctricos a lo largo de los años del modelo. Esta evolución positiva puede estar relacionada con avances tecnológicos en la industria de vehículos eléctricos, como mejoras en la eficiencia de las baterías y el desarrollo de nuevas tecnologías de propulsión



PRIMERA EVALUACIÓN DEL SEGUNDO BIMESTRE ANÁLISIS DE DATOS



eléctrica. Según un artículo de Real Python, los avances en la tecnología de baterías han sido un factor clave en el aumento del rango eléctrico de los vehículos, lo que respalda la tendencia observada en el gráfico de líneas.

- ✓ Impacto del Avance Tecnológico en la Industria Automotriz: La representación visual del rango eléctrico por año del modelo en el gráfico de líneas permite apreciar el impacto del avance tecnológico en la industria automotriz. Este tipo de análisis es fundamental para comprender cómo la evolución tecnológica ha influido en las capacidades de los vehículos eléctricos a lo largo del tiempo. Según un tutorial de Real Python, el análisis de datos y la visualización son herramientas clave para comprender y comunicar el impacto de la tecnología en diversas industrias, incluida la automotriz.

Las conclusiones fueron en base a través de datos leídos y comparados a través de estos sitios web:

<https://energy5.com/es/avances-tecnologicos-en-el-mantenimiento-de-vehiculos-electricos>

<https://energy5.com/es/los-avances-tecnologicos-hacen-que-el-mantenimiento-de-vehiculos-electricos-sea-mas-facil-que-nunca>

Entregables:

Archivo pdf con el link del repositorio de github que contenga el archivo ipynb.

Link del GitHub:

https://github.com/CarlosAUD/Prueba_AnalisisDeDatos_2Bimestre.git

Tiempo:

2 horas